The Laws of Large Numbers

Manuel Eberl

Abstract

The Law of Large Numbers states that, informally, if one performs a random experiment X many times and takes the average of the results, that average will be very close to the expected value $E[X]$.

More formally, let $(X_i)_{i\in\mathbb{N}}$ be a sequence of independently identically distributed random variables whose expected value $E[X_1]$ exists. Denote the running average of $X_1, \ldots, X_n$ for $\overline{X}_n$. Then:

- The Weak Law of Large Numbers states that $\overline{X}_n \longrightarrow E[X_1]$ in probability for $n \to \infty$, i.e. $\mathcal{P}(|\overline{X}_n - E[X_1]| > \varepsilon) \longrightarrow 0$ for $n \to \infty$ for any $\varepsilon > 0$.
- The Strong Law of Large Numbers states that $\overline{X}_n \longrightarrow E[X_1]$ almost surely for $n \to \infty$, i.e. $\mathcal{P}(\overline{X}_n \longrightarrow E[X_1]) = 1$.

In this entry, I formally prove the strong law and from it the weak law. The approach used for the proof of the strong law is a particularly quick and slick one based on ergodic theory, which was formalised by Gouëzel in another AFP entry.

Contents

1 The Laws of Large Numbers

theory *Laws-of-Large-Numbers*
  imports *Ergodic-Theory.Shift-Operator*
begin

We prove the strong law of large numbers in the following form: Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables over a probability space M. Further assume that the expected value $E[X_0]$ of X_0 exists. Then the sequence of random variables

$$\overline{X}_n = \frac{1}{n}\sum_{i=0}^{n} X_i$$

of running averages almost surely converges to $E[X_0]$. This means that

$$\mathcal{P}[\overline{X}_n \longrightarrow E[X_0]] = 1 \ .$$

We start with the strong law.

1.1 The strong law

The proof uses Birkhoff's Theorem from Gouëzel's formalisation of ergodic theory [1] and the fact that the shift operator $T(x_1, x_2, x_3, \ldots) = (x_2, x_3, \ldots)$ is ergodic. This proof can be found in various textbooks on probability theory/ergodic theory, e.g. the ones by Krengel [2, p. 24] and Simmonet [3, Chapter 15, pp. 311–325].

theorem (**in** *prob-space*) *strong-law-of-large-numbers-iid*:
  fixes X :: *nat* ⇒ *'a* ⇒ *real*
  assumes *indep*: *indep-vars* (λ-. *borel*) *X UNIV*
  assumes *distr*: $\bigwedge$*i*. *distr M borel* (*X i*) = *distr M borel* (*X 0*)
  assumes *L1*: *integrable M* (*X 0*)
  shows *AE x in M*. (λ*n*. ($\sum$*i*<*n*. *X i x*) / *n*) ⟶ *expectation* (*X 0*)
⟨*proof*⟩

1.2 The weak law

To go from the strong law to the weak one, we need the fact that almost sure convergence implies convergence in probability. We prove this for sequences of random variables here.

lemma (**in** *prob-space*) *AE-convergence-imp-convergence-in-prob*:
  assumes [*measurable*]: $\bigwedge$*i*. *random-variable borel* (*X i*) *random-variable borel Y*
  assumes *AE*: *AE x in M*. (λ*i*. *X i x*) ⟶ *Y x*
  assumes ε > (*0* :: *real*)
  shows (λ*i*. *prob* {*x*∈*space M*. |*X i x* − *Y x*| > ε}) ⟶ *0*
⟨*proof*⟩

The weak law is now a simple corollary: we again have the same setting as before. The weak law now states that $\overline{X}_n$ converges to $E[X_0]$ in probability. This means that for any $\varepsilon > 0$, the probability that $|\overline{X}_n - X_0| > \varepsilon$ vanishes as $n \to \infty$.

corollary (**in** *prob-space*) *weak-law-of-large-numbers-iid*:
  fixes X :: *nat* ⇒ *'a* ⇒ *real* **and** ε :: *real*
  assumes *indep*: *indep-vars* (λ-. *borel*) *X UNIV*
  assumes *distr*: $\bigwedge$*i*. *distr M borel* (*X i*) = *distr M borel* (*X 0*)
  assumes *L1*: *integrable M* (*X 0*)
  assumes ε > *0*
  shows (λ*n*. *prob* {*x*∈*space M*. |($\sum$*i*<*n*. *X i x*) / *n* − *expectation* (*X 0*)| > ε})
⟶ *0*
⟨*proof*⟩

end

1.3 Example

theory *Laws-of-Large-Numbers-Example*
  imports *Laws-of-Large-Numbers*
begin

As an example, we apply the strong law to the proportion of successes in an independent sequence of coin flips with success probability p. We will show that proportion of successful coin flips among the first n attempts almost surely converges to p as $n \to \infty$.

lemma (**in** *prob-space*) *indep-vars-iff-distr-eq-PiM'*:
  fixes I :: *'i set* **and** X :: *'i ⇒ 'a ⇒ 'b*
  assumes $I \neq \{\}$
  assumes *rv*: $\bigwedge i.\ i \in I \Longrightarrow$ *random-variable* $(M'\ i)\ (X\ i)$
  shows *indep-vars* $M'\ X\ I \longleftrightarrow$
    distr $M\ (\Pi_M\ i{\in}I.\ M'\ i)\ (\lambda x.\ \lambda i{\in}I.\ X\ i\ x) = (\Pi_M\ i{\in}I.$ *distr* $M\ (M'\ i)$ $(X\ i))$
⟨*proof*⟩

lemma *indep-vars-PiM-components*:
  assumes $\bigwedge i.\ i \in A \Longrightarrow$ *prob-space* $(M\ i)$
  shows *prob-space.indep-vars* $(PiM\ A\ M)\ M\ (\lambda i\ f.\ f\ i)\ A$
⟨*proof*⟩

lemma *indep-vars-PiM-components'*:
  assumes $\bigwedge i.\ i \in A \Longrightarrow$ *prob-space* $(M\ i)$
  assumes $\bigwedge i.\ i \in A \Longrightarrow g\ i \in M\ i \to_M N\ i$
  shows *prob-space.indep-vars* $(PiM\ A\ M)\ N\ (\lambda i\ f.\ g\ i\ (f\ i))\ A$
  ⟨*proof*⟩

lemma *integrable-bernoulli-pmf* [*intro*]:
  fixes f :: *bool ⇒ 'a* :: {*banach*, *second-countable-topology*}
  shows *integrable* (*bernoulli-pmf* p) f
  ⟨*proof*⟩

lemma *expectation-bernoulli-pmf*:
  fixes f :: *bool ⇒ 'a* :: {*banach*, *second-countable-topology*}
  assumes *p*: $p \in \{0..1\}$
  shows *measure-pmf.expectation* (*bernoulli-pmf* p) $f = p *_R f$ *True* $+ (1 - p)$ $*_R f$ *False*
  ⟨*proof*⟩

experiment
  fixes *p* :: *real*
  assumes *p*: $p \in \{0..1\}$
begin

definition *M* :: (*nat* ⇒ *bool*) *measure*
  where $M = (\Pi_M\ i{\in}(UNIV :: nat\ set).\ measure\text{-}pmf\ (bernoulli\text{-}pmf\ p))$

definition *X* :: *nat* ⇒ (*nat* ⇒ *bool*) ⇒ *real*
  where $X = (\lambda i\ f.\ if\ f\ i\ then\ 1\ else\ 0)$

interpretation *prob-space M*
  ⟨*proof*⟩

lemma *random-variable-component*: *random-variable* (*count-space UNIV*) ($\lambda f.\ f\ i$)
  ⟨*proof*⟩

lemma *random-variable-X* [*measurable*]: *random-variable borel* (*X i*)
  ⟨*proof*⟩

lemma *distr-M-component*: *distr M* (*count-space UNIV*) ($\lambda f.\ f\ i$) = *measure-pmf* (*bernoulli-pmf p*)
⟨*proof*⟩

lemma *distr-M-X*:
  distr M borel (*X i*) = *distr* (*measure-pmf* (*bernoulli-pmf p*)) *borel* ($\lambda b.\ if\ b\ then\ 1\ else\ 0$)
⟨*proof*⟩

lemma *X-has-expectation*: *integrable M* (*X 0*)
⟨*proof*⟩

lemma *indep*: *indep-vars* (λ-. *borel*) *X UNIV*
  ⟨*proof*⟩

lemma *expectation-X*: *expectation* (*X i*) = *p*
⟨*proof*⟩

theorem *AE f in M*. ($\lambda n.\ card\ \{i.\ i < n \wedge f\ i\}\ /\ n) \longrightarrow p$
⟨*proof*⟩

end

end

Acknowledgements. I thank Sébastien Gouëzel for providing advice and context about the law of large numbers and ergodic theory. I do not actually

know any ergodic theory and without him, I would probably have shied away from formalising this.

References

[1] Sébastien Gouëzel. Ergodic theory. *Archive of Formal Proofs*, December 2015. ISSN 2150-914x. https://isa-afp.org/entries/Ergodic_Theory.html, Formal proof development.

[2] Ulrich Krengel. *Ergodic Theorems*. De Gruyter, January 1985. doi:10.1515/9783110844641.

[3] Michel Simonnet. *Measures and Probabilities*. Springer New York, New York, NY, 1996. ISBN 978-1-4612-4012-9. doi:10.1007/978-1-4612-4012-9_15.